Compositional and lexical semantics

- Compositional semantics: the construction of meaning (generally expressed as logic) based on syntax.
  This lecture:
  - Semantics with FS grammars
- Lexical semantics: the meaning of individual words.
  This lecture:
  - lexical semantic relations and WordNet
  - one technique for word sense disambiguation

Simple compositional semantics in feature structures

- Semantics is built up along with syntax
- Subcategorization 'slot' filling instantiates syntax
- Formally equivalent to logical representations (below: predicate calculus with no quantifiers)
- Alternative FS encodings possible

Objective: obtain the following semantics for *they like fish*:

$\text{pron}(x) \wedge (\text{like\_v}(x, y) \wedge \text{fish\_n}(y))$

Feature structure encoding:

$$\begin{bmatrix} \text{PRED} & \textbf{and} \\ \text{ARG1} & \begin{bmatrix} \text{PRED} & \textbf{pron} \\ \text{ARG1} & \boxed{1} \end{bmatrix} \\ \text{ARG2} & \begin{bmatrix} \text{PRED} & \textbf{and} \\ \text{ARG1} & \begin{bmatrix} \text{PRED} & \textbf{like\_v} \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2} \end{bmatrix} \\ \text{ARG2} & \begin{bmatrix} \text{PRED} & \textbf{fish\_n} \\ \text{ARG1} & \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Noun entry

$$\text{fish} \quad \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{CAT} & \textbf{noun} \\ \text{AGR} & \end{bmatrix} \\ \text{COMP} & \textbf{filled} \\ \text{SPR} & \textbf{filled} \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & \boxed{1} \\ \text{PRED} & \textbf{fish\_n} \\ \text{ARG1} & \boxed{1} \end{bmatrix} \end{bmatrix}$$

- Corresponds to fish(x) where the INDEX points to the characteristic variable of the noun (that is x).
  The INDEX is unambiguous here, but e.g., picture$(x, y) \wedge$ sheep(y)
  picture of sheep

Verb entry

$$\text{like} \quad \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{CAT} & \textbf{verb} \\ \text{AGR} & \textbf{pl} \end{bmatrix} \\ \text{COMP} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{CAT} & \textbf{noun} \end{bmatrix} \\ \text{COMP} & \textbf{filled} \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & \boxed{2} \end{bmatrix} \end{bmatrix} \\ \text{SPR} & \begin{bmatrix} \text{HEAD} & \begin{bmatrix} \text{CAT} & \textbf{noun} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{INDEX} & \boxed{1} \end{bmatrix} \end{bmatrix} \\ \text{SEM} & \begin{bmatrix} \text{PRED} & \textbf{like\_v} \\ \text{ARG1} & \boxed{1} \\ \text{ARG2} & \boxed{2} \end{bmatrix} \end{bmatrix}$$

- Linking syntax and semantics: ARG1 linked to the SPR's index; ARG2 linked to the COMP index.

COMP filling rule

$$
\begin{bmatrix}
\text{HEAD } \boxed{1} \\
\text{COMP } \textbf{filled} \\
\text{SPR } \boxed{3} \\
\text{SEM } \begin{bmatrix} \text{PRED } \textbf{and} \\ \text{ARG1 } \boxed{4} \\ \text{ARG2 } \boxed{5} \end{bmatrix}
\end{bmatrix}
\rightarrow
\begin{bmatrix}
\text{HEAD } \boxed{1} \\
\text{COMP } \boxed{2} \\
\text{SPR } \boxed{3} \\
\text{SEM } \boxed{4}
\end{bmatrix},\ \boxed{2}\begin{bmatrix}
\text{COMP } \textbf{filled} \\
\text{SEM } \boxed{5}
\end{bmatrix}
$$

- As last time: object of the verb (DTR2) ‘fills’ the COMP slot
- New: semantics on the mother is the ‘and’ of the semantics of the dtrs

Apply to *like*

$$
\begin{bmatrix}
\text{HEAD } \boxed{1} \\
\text{COMP } \textbf{filled} \\
\text{SPR } \boxed{3} \\
\text{SEM } \begin{bmatrix} \text{PRED } \textbf{and} \\ \text{ARG1 } \boxed{4} \\ \text{ARG2 } \boxed{5} \end{bmatrix}
\end{bmatrix}
\rightarrow
$$

$$
\begin{bmatrix}
\text{HEAD } \boxed{1} \begin{bmatrix} \text{CAT } \textbf{verb} \\ \text{AGR } \textbf{pl} \end{bmatrix} \\
\text{COMP } \boxed{2} \begin{bmatrix} \text{HEAD } \begin{bmatrix} \text{CAT } \textbf{noun} \end{bmatrix} \\ \text{COMP } \textbf{filled} \\ \text{SEM } \boxed{5} \begin{bmatrix} \text{INDEX } \boxed{6} \end{bmatrix} \end{bmatrix} \\
\text{SPR } \boxed{3} \begin{bmatrix} \text{HEAD } \begin{bmatrix} \text{CAT } \textbf{noun} \end{bmatrix} \\ \text{SEM } \begin{bmatrix} \text{INDEX } \boxed{7} \end{bmatrix} \end{bmatrix} \\
\text{SEM } \boxed{4} \begin{bmatrix} \text{PRED } \textbf{like\_v} \\ \text{ARG1 } \boxed{7} \\ \text{ARG2 } \boxed{6} \end{bmatrix}
\end{bmatrix}, \ \boxed{2}
$$

Apply to *like fish*

$$\begin{bmatrix} \text{HEAD } \boxed{1} \\ \text{COMP } \textbf{filled} \\ \text{SPR } \boxed{3} \\ \text{SEM } \begin{bmatrix} \text{PRED } \textbf{and} \\ \text{ARG1 } \boxed{4} \\ \text{ARG2 } \boxed{5} \end{bmatrix} \end{bmatrix} \rightarrow$$

$$\begin{bmatrix} \text{HEAD } \boxed{1} \begin{bmatrix} \text{CAT } \textbf{verb} \\ \text{AGR } \textbf{pl} \end{bmatrix} \\ \text{COMP } \boxed{2} \begin{bmatrix} \text{HEAD } \begin{bmatrix} \text{CAT } \textbf{noun} \\ \text{AGR} \end{bmatrix} \\ \text{COMP } \textbf{filled} \\ \text{SPR } \textbf{filled} \\ \text{SEM } \boxed{5} \begin{bmatrix} \text{INDEX } \boxed{6} \\ \text{PRED } \textbf{fish\_n} \\ \text{ARG1 } \boxed{6} \end{bmatrix} \end{bmatrix} \\ \text{SPR } \boxed{3} \begin{bmatrix} \text{HEAD } \begin{bmatrix} \text{CAT } \textbf{noun} \end{bmatrix} \\ \text{SEM } \begin{bmatrix} \text{INDEX } \boxed{7} \end{bmatrix} \end{bmatrix} \\ \text{SEM } \boxed{4} \begin{bmatrix} \text{PRED } \textbf{like\_v} \\ \text{ARG1 } \boxed{7} \\ \text{ARG2 } \boxed{6} \end{bmatrix} \end{bmatrix}, \boxed{2}$$

$$\text{like\_v}(x, y) \wedge \text{fish\_n}(y)$$

Logic in semantic representation

- Meaning representation for a sentence is called the *logical form*
- Standard approach to composition in theoretical linguistics is lambda calculus, building FOPC or higher order representation
- Representation above is impoverished but can build FOPC in FSs
- Theorem proving
- Generation: starting point is logical form, not string.

Meaning postulates

- e.g.,
  $\forall x[\text{bachelor}'(x) \rightarrow \text{man}'(x) \wedge \text{unmarried}'(x)]$
- usable with compositional semantics and theorem provers
- e.g. from 'Kim is a bachelor', we can construct the LF

  $$\text{bachelor}'(\text{Kim})$$

  and then deduce

  $$\text{unmarried}'(\text{Kim})$$

- OK for narrow domains, but 'classical' lexical semantic relations are more generally useful

Lexical semantic relations

Hyponymy: IS-A:

- (a sense of) *dog* is a *hyponym* of (a sense of) *animal*
- *animal* is a *hypernym* of *dog*
- hyponymy relationships form a *taxonomy*
- works best for concrete nouns

Meronomy: PART-OF e.g., *arm* is a *meronym* of *body*, *steering wheel* is a meronym of *car* (piece vs part)

Synonymy e.g., *aubergine/eggplant*

Antonymy e.g., *big/little*

WordNet

- large scale, open source resource for English
- hand-constructed
- wordnets being built for other languages
- organized into *synsets*: synonym sets (near-synonyms)

Overview of adj red:

```
1. (43) red, reddish, ruddy,
blood-red, carmine, cerise,
cherry, cherry-red, crimson,
ruby, ruby-red, scarlet --
(having any of numerous bright
or strong colors reminiscent
of the color of blood or
cherries or tomatoes or rubies)
2. (8) red, reddish --
((used of hair or fur) of a
reddish brown color; "red deer";
reddish hair")
```

Hyponymy in WordNet

```
Sense 6
big cat, cat
       => leopard, Panthera pardus
           => leopardess
           => panther
       => snow leopard, ounce, Panthera uncia
       => jaguar, panther, Panthera onca,
                                    Felis onca
       => lion, king of beasts, Panthera leo
           => lioness
           => lionet
       => tiger, Panthera tigris
           => Bengal tiger
           => tigress
       => liger
       => tiglon, tigon
       => cheetah, chetah, Acinonyx jubatus
       => saber-toothed tiger, sabertooth
           => Smiledon californicus
           => false saber-toothed tiger
```

Some uses of lexical semantics

- Semantic classification: e.g., for selectional restrictions (e.g., the object of *eat* has to be something edible) and for named entity recognition
- Shallow inference: ‘X murdered Y’ implies ‘X killed Y’ etc
- Back-off to semantic classes in some statistical approaches
- Word-sense disambiguation
- Machine Translation: if you can’t translate a term, substitute a hypernym
- Query expansion: if a search doesn’t return enough results, one option is to replace an over-specific term with a hypernym

Word sense disambiguation

Needed for many applications, problematic for large domains. May depend on:

- frequency: e.g., *diet*: the food sense (or senses) is much more frequent than the parliament sense (Diet of Wurms)
- collocations: e.g. *striped bass* (the fish) vs *bass guitar*: syntactically related or in a window of words (latter sometimes called ‘cooccurrence’). Generally ‘one sense per collocation’.
- selectional restrictions/preferences (e.g., *Kim eats bass*, must refer to fish

WSD techniques

- supervised learning: cf. POS tagging from lecture 3. But sense-tagged corpora are difficult to construct, algorithms need far more data than POS tagging
- unsupervised learning (see below)
- Machine readable dictionaries (MRDs)
- selectional preferences: don't work very well by themselves, useful in combination with other techniques

WSD by (almost) unsupervised learning

Disambiguating *plant* (factory vs vegetation senses):

1. Find contexts in training corpus:

sense	training example
?	company said that the *plant* is still operating
?	although thousands of *plant* and animal species
?	zonal distribution of *plant* life
?	company manufacturing *plant* is in Orlando
	etc

2. Identify some seeds to disambiguate a few uses. e.g., 'plant life' for vegetation use (A) 'manufacturing plant' for factory use (B):

sense	training example
?	company said that the *plant* is still operating
?	although thousands of *plant* and animal species
A	zonal distribution of *plant* life
B	company manufacturing *plant* is in Orlando
	etc

3. Train a *decision list* classifier on the Sense A/Sense B examples.

reliability	criterion	sense
8.10	*plant* life	A
7.58	manufacturing *plant*	B
6.27	*animal* within 10 words of *plant*	A
	etc	

4. Apply the classifier to the training set and add reliable examples to A and B sets.

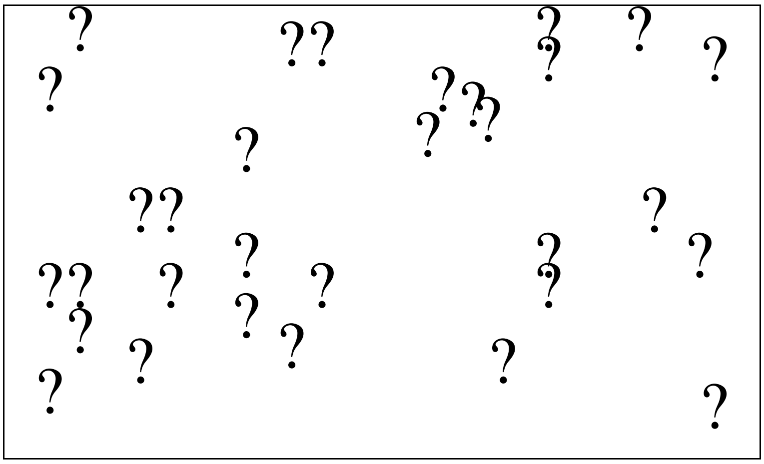
sense	training example
?	company said that the *plant* is still operating
A	although thousands of *plant* and animal species
A	zonal distribution of *plant* life
B	company manufacturing *plant* is in Orlando
	etc

5. Iterate the previous steps 3 and 4 until convergence
6. Apply the classifier to the unseen test data

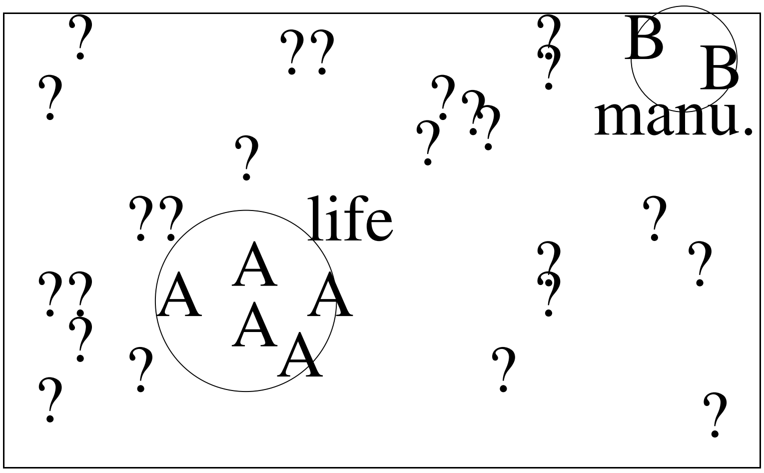
‘one sense per discourse’: can be used as an additional refinement

Yarowsky (1995): schematically

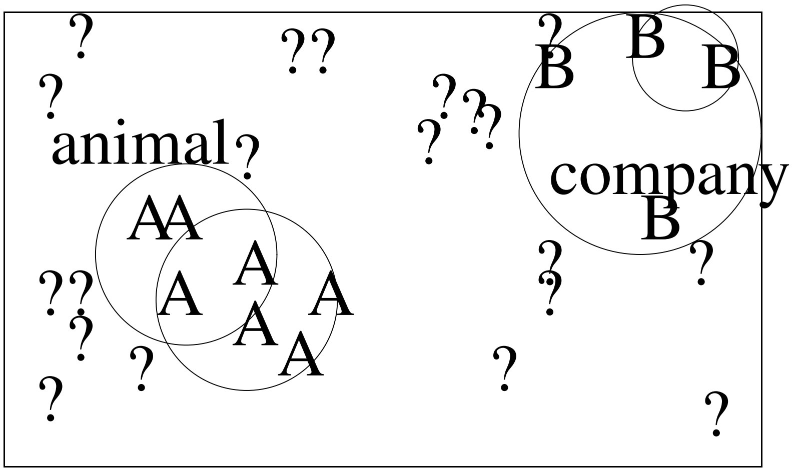
Initial state



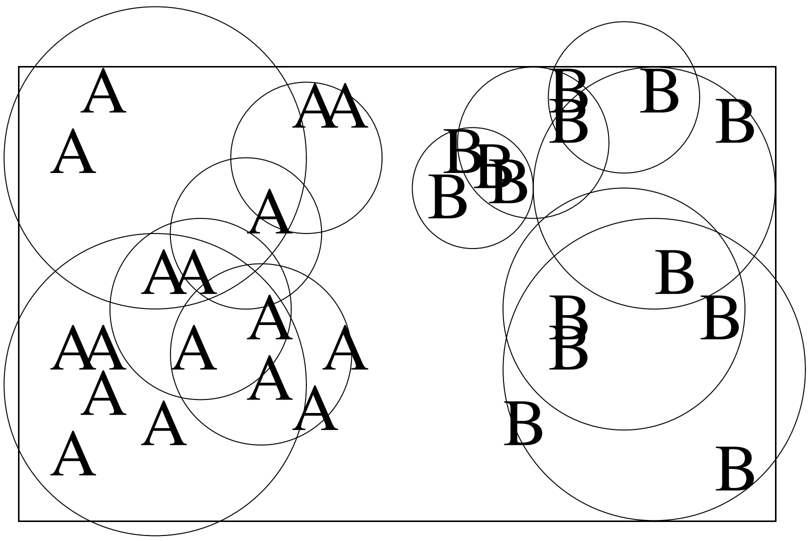
Seeds



Iterating:



Final:



Evaluation of WSD

- SENSEVAL and SENSEVAL-2 competitions
- evaluate against WordNet
- baseline: pick most frequent sense — hard to beat (but don't always know most frequent sense)
- human ceiling varies with words
- MT task: more objective but sometimes doesn't correspond to polysemy in source language